12: Bootstrap Confidence Intervals

Stat 120 | Fall 2025

Prof Amanda Luby

1 Part 1: StatKey

Since we didn't get to the StatKey activity during last class, start with completing one problem from the activity (problems assigned by group below). Once your group feels good about your answer, Groups 1-4 should add their solutions in the collaborative key google doc and Groups 5-8 are responsible for checking answers and providing alternative solutions for any discrepancies.

	Responsible for adding answers	Responsible for checking answers
Problem 1: Atlanta Commute Times	Group 1	Group 5
Problem 2: Compassionate Rats	Group 2	Group 6
Problem 3: Do Teen Problems Differ Based on Income Level?	Group 3	Group 7
Problem 4: Atlanta Commute Times (again)	Group 4	Group 8

2 Part 2: R

```
library(tidyverse)
library(broom)
library(patchwork)
library(CarletonStats)
```

The data set Pew.csv contains part of a survey conducted by the Pew Research Center in January 2014. One of the questions they asked was, "Overall, when you add up all the advantages and disadvantages of the internet, would you say that the internet has been mostly a good thing, a bad thing, or some of both?"

The variable `values` codes the response as "good" if the respondent said the internet has been a good thing and "bad" otherwise (this includes "a bad thing" and "some of both").

Let's see if this differs based on whether the respondent is 50 years or older or not.

```
Pew <- read.csv("http://math.carleton.edu/Stat120/RLabManual/Pew.csv")

Pew <- Pew |>
  mutate(
    over = ifelse(age >=50, "over50", "under50"),
    over2 = ifelse(age >=50, 1, 0)
  )
```

1. Create an appropriate visualization of `values` conditioned on `over`. Does it appear that the proportion of over 50 year olds that said the internet has been a good thing is approximately the same as the under 50's?

2. Compute the exact proportions for each age group.

3. In order to use the `boot()` command, the response must be coded as a binary variable of 0's and 1's (instead of the names of the categories). Pull up the "spreadsheet view" of the data and confirm that `values` and `values2` contain the same information stored in a different form.

Means vs Proportions?

Since the values are 0's and 1's, when you compute the mean, you will be computing an expression of the form

$$\frac{1+1+0+1+0+...+0}{n}$$

which is equivalent to the proportion of 1's. This means that we can treat proportions as means. Yay!

4. Create a bootstrap distribution of the difference in proportions and give a sentence interpreting the 95% percentile interval. Use the variable `values2`, which is behavior recoded so that 1 = "good", 0 = "bad".

5. Compute the 95% confidence interval based on the 95% rule. How does it compare to the percentile interval?

When you are done...

Submit your answers to 1-5 on a PDF on gradescope. You should submit one PDF per group and *make sure that all group members who were in class are listed on the submission!*